

L12: Entry 266 of 288

File: USPT

Jul 6, 1999

US-PAT-NO: 5919665

DOCUMENT-IDENTIFIER: US 5919665 A

TITLE: Vaccine for clostridium botulinum neurotoxin

DATE-ISSUED: July 6, 1999

INVENTOR-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY
Williams; James A.	Madison	WI		

US-CL-CURRENT: 435/71.1, 435/252.3, 435/320.1, 530/350, 530/825, 536/23.4

CLAIMS:

I claim:

1. A soluble fusion protein comprising a non-toxin protein sequence and a portion of the Clostridium botulinum type A toxin, said portion of the Clostridium botulinum type A toxin comprising a portion of the sequence of SEQ ID NO:28.

2. The fusion protein of claim 1, wherein said portion of the Clostridium botulinum type A toxin sequence comprises SEQ ID NO:23.

3. The fusion protein of claim 1, wherein said non-toxin protein sequence comprises a poly-histidine tract.

4. The fusion protein of claim 3, which comprises SEQ ID NO:26.

5. The fusion protein of claim 1, wherein said fusion protein is substantially endotoxin-free.

6. A host cell containing a recombinant expression vector, said vector encoding a protein comprising at least a portion of a Clostridium botulinum type A toxin protein sequence of SEQ ID NO:28, and wherein said host cell is capable of expressing said protein as a soluble protein in said host cell at a level greater than or equal to 0.75% of the total cellular protein.

7. The host cell of claim 6, wherein said portion of a toxin comprises SEQ ID NO:23.

8. The host cell of claim 6, wherein said fusion protein comprises SEQ ID NO:26.

9. The host cell of claim 6, wherein said host cell is capable of expressing said protein in said host cell at a level greater than or equal to 20% of the total cellular protein.

10. A soluble fusion protein, comprising at least a portion of Clostridium botulinum C fragment linked to a poly-histidine tag.

Example 22, 23, 24,

7/8/25
E. al
w
C
tag

CS tag
His tagged

-continued

Arg Val Tyr Ile Asn Val Val Val Lys Asn Lys Glu Tyr Arg Leu Ala	
325 330 335	
ACC AAT GCT TCT CAG GCT GGT GTA GAA AAG ATC TTG TCT GCT CTG GAA	1056
Thr Asn Ala Ser Gln Ala Gly Val Glu Lys Ile Leu Ser Ala Leu Glu	
340 345 350	
ATC CCG GAC GTT GGT AAT CTG TCT CAG GTA GTT GTA ATG AAA TCC AAG	1104
Ile Pro Asp Val Gly Asn Leu Ser Gln Val Val Val Met Lys Ser Lys	
355 360 365	
AAC GAC CAG GGT ATC ACT AAC AAA TGC AAA ATG AAT CTG CAG GAC AAC	1152
Asn Asp Gln Gly Ile Thr Asn Lys Cys Lys Met Asn Leu Gln Asp Asn	
370 375 380	
AAT GGT AAC GAT ATC GGT TTC ATC GGT TTC CAC CAG TTC AAC AAT ATC	1200
Asn Gly Asn Asp Ile Gly Phe Ile Gly Phe His Gln Phe Asn Asn Ile	
385 390 395 400	
GCT AAA CTG GTT GCT TCC AAC TGG TAC AAT CGT CAG ATC GAA CGT TCC	1248
Ala Lys Leu Val Ala Ser Asn Trp Tyr Asn Arg Gln Ile Glu Arg Ser	
405 410 415	
TCT CGC ACT CTG GGT TGC TCT TGG GAG TTC ATC CCG GTT GAT GAC GGT	1296
Ser Arg Thr Leu Gly Cys Ser Trp Glu Phe Ile Pro Val Asp Asp Gly	
420 425 430	
TGG GGT GAA CGT CCG CTG TAACCCGGGA AAGCTT	1330
Trp Gly Glu Arg Pro Leu	
435	

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 438 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

Met Ala Arg Leu Leu Ser Thr Phe Thr Glu Tyr Ile Lys Asn Ile Ile	
1 5 10 15	
Asn Thr Ser Ile Leu Asn Leu Arg Tyr Glu Ser Asn His Leu Ile Asp	
20 25 30	
Leu Ser Arg Tyr Ala Ser Lys Ile Asn Ile Gly Ser Lys Val Asn Phe	
35 40 45	
Asp Pro Ile Asp Lys Asn Gln Ile Gln Leu Phe Asn Leu Glu Ser Ser	
50 55 60	
Lys Ile Glu Val Ile Leu Lys Asn Ala Ile Val Tyr Asn Ser Met Tyr	
65 70 75 80	
Glu Asn Phe Ser Thr Ser Phe Trp Ile Arg Ile Pro Lys Tyr Phe Asn	
85 90 95	
Ser Ile Ser Leu Asn Asn Glu Tyr Thr Ile Ile Asn Cys Met Glu Asn	
100 105 110	
Asn Ser Gly Trp Lys Val Ser Leu Asn Tyr Gly Glu Ile Ile Trp Thr	
115 120 125	
Leu Gln Asp Thr Gln Glu Ile Lys Gln Arg Val Val Phe Lys Tyr Ser	
130 135 140	
Gln Met Ile Asn Ile Ser Asp Tyr Ile Asn Arg Trp Ile Phe Val Thr	
145 150 155 160	
Ile Thr Asn Asn Arg Leu Asn Asn Ser Lys Ile Tyr Ile Asn Gly Arg	
165 170 175	
Leu Ile Asp Gln Lys Pro Ile Ser Asn Leu Gly Asn Ile His Ala Ser	
180 185 190	
Asn Asn Ile Met Phe Lys Leu Asp Gly Cys Arg Asp Thr His Arg Tyr	

-continued

195	200	205
Ile Trp Ile Lys Tyr Phe Asn Leu Phe Asp Lys Glu Leu Asn Glu Lys 210 215 220		
Glu Ile Lys Asp Leu Tyr Asp Asn Gln Ser Asn Ser Gly Ile Leu Lys 225 230 235 240		
Asp Phe Trp Gly Asp Tyr Leu Gln Tyr Asp Lys Pro Tyr Tyr Met Leu 245 250 255		
Asn Leu Tyr Asp Pro Asn Lys Tyr Val Asp Val Asn Asn Val Gly Ile 260 265 270		
Arg Gly Tyr Met Tyr Leu Lys Gly Pro Arg Gly Ser Val Met Thr Thr 275 280 285		
Asn Ile Tyr Leu Asn Ser Ser Leu Tyr Arg Gly Thr Lys Phe Ile Ile 290 295 300		
Lys Lys Tyr Ala Ser Gly Asn Lys Asp Asn Ile Val Arg Asn Asn Asp 305 310 315 320		
Arg Val Tyr Ile Asn Val Val Val Lys Asn Lys Glu Tyr Arg Leu Ala 325 330 335		
Thr Asn Ala Ser Gln Ala Gly Val Glu Lys Ile Leu Ser Ala Leu Glu 340 345 350		
Ile Pro Asp Val Gly Asn Leu Ser Gln Val Val Val Met Lys Ser Lys 355 360 365		
Asn Asp Gln Gly Ile Thr Asn Lys Cys Lys Met Asn Leu Gln Asp Asn 370 375 380		
Asn Gly Asn Asp Ile Gly Phe Ile Gly Phe His Gln Phe Asn Asn Ile 385 390 395 400		
Ala Lys Leu Val Ala Ser Asn Trp Tyr Asn Arg Gln Ile Glu Arg Ser 405 410 415		
Ser Arg Thr Leu Gly Cys Ser Trp Glu Phe Ile Pro Val Asp Asp Gly 420 425 430		
Trp Gly Glu Arg Pro Leu 435		

(2) INFORMATION FOR SEQ ID NO:24:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 23 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: unknown
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

Met Gly His His His His His His His His His Ser Ser Gly His
 1 5 10 15

Ile Glu Gly Arg His Met Ala
 20

(2) INFORMATION FOR SEQ ID NO:25:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1402 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

- (A) NAME/KEY: CDS
 (B) LOCATION: 1..1386

-continued

Pro	Arg	Gly	Ser	Val	Met	Thr	Thr	Asn	Ile	Tyr	Leu	Asn	Ser	Ser	Leu	
305					310					315					320	
TAC	CGT	GGT	ACC	AAA	TTC	ATC	ATC	AAG	AAA	TAC	GCG	TCT	GGT	AAC	AAG	1008
Tyr	Arg	Gly	Thr	Lys	Phe	Ile	Ile	Lys	Lys	Tyr	Ala	Ser	Gly	Asn	Lys	
				325				330					335			
GAC	AAT	ATC	GTT	CGC	AAC	AAT	GAT	CGT	GTA	TAC	ATC	AAT	GTT	GTA	GTT	1056
Asp	Asn	Ile	Val	Arg	Asn	Asn	Asp	Arg	Val	Tyr	Ile	Asn	Val	Val	Val	
			340					345					350			
AAG	AAC	AAA	GAA	TAC	CGT	CTG	GCT	ACC	AAT	GCT	TCT	CAG	GCT	GGT	GTA	1104
Lys	Asn	Lys	Glu	Tyr	Arg	Leu	Ala	Thr	Asn	Ala	Ser	Gln	Ala	Gly	Val	
		355				360						365				
GAA	AAG	ATC	TTG	TCT	GCT	CTG	GAA	ATC	CCG	GAC	GTT	GGT	AAT	CTG	TCT	1152
Glu	Lys	Ile	Leu	Ser	Ala	Leu	Glu	Ile	Pro	Asp	Val	Gly	Asn	Leu	Ser	
	370					375					380					
CAG	GTA	GTT	GTA	ATG	AAA	TCC	AAG	AAC	GAC	CAG	GGT	ATC	ACT	AAC	AAA	1200
Gln	Val	Val	Val	Met	Lys	Ser	Lys	Asn	Asp	Gln	Gly	Ile	Thr	Asn	Lys	
	385				390					395					400	
TGC	AAA	ATG	AAT	CTG	CAG	GAC	AAC	AAT	GGT	AAC	GAT	ATC	GGT	TTC	ATC	1248
Cys	Lys	Met	Asn	Leu	Gln	Asp	Asn	Asn	Gly	Asn	Asp	Ile	Gly	Phe	Ile	
			405						410					415		
GGT	TTC	CAC	CAG	TTC	AAC	AAT	ATC	GCT	AAA	CTG	GTT	GCT	TCC	AAC	TGG	1296
Gly	Phe	His	Gln	Phe	Asn	Asn	Ile	Ala	Lys	Leu	Val	Ala	Ser	Asn	Trp	
			420					425					430			
TAC	AAT	CGT	CAG	ATC	GAA	CGT	TCC	TCT	CGC	ACT	CTG	GGT	TGC	TCT	TGG	1344
Tyr	Asn	Arg	Gln	Ile	Glu	Arg	Ser	Ser	Arg	Thr	Leu	Gly	Cys	Ser	Trp	
		435					440					445				
GAG	TTC	ATC	CCG	GTT	GAT	GAC	GGT	TGG	GGT	GAA	CGT	CCG	CTG			1386
Glu	Phe	Ile	Pro	Val	Asp	Asp	Gly	Trp	Gly	Glu	Arg	Pro	Leu			
	450					455				460						
TAACCCGGGA AAGCTT																1402

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 462 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

Met Gly His His His His His His His His His Ser Ser Gly His
 1 5 10 15

Ile Glu Gly Arg His Met Ala Ser Met Ala Arg Leu Leu Ser Thr Phe
 20 25 30

Thr Glu Tyr Ile Lys Asn Ile Ile Asn Thr Ser Ile Leu Asn Leu Arg
 35 40 45

Tyr Glu Ser Asn His Leu Ile Asp Leu Ser Arg Tyr Ala Ser Lys Ile
 50 55 60

Asn Ile Gly Ser Lys Val Asn Phe Asp Pro Ile Asp Lys Asn Gln Ile
 65 70 75 80

Gln Leu Phe Asn Leu Glu Ser Ser Lys Ile Glu Val Ile Leu Lys Asn
 85 90 95

Ala Ile Val Tyr Asn Ser Met Tyr Glu Asn Phe Ser Thr Ser Phe Trp
 100 105 110

Ile Arg Ile Pro Lys Tyr Phe Asn Ser Ile Ser Leu Asn Asn Glu Tyr
 115 120 125

Thr Ile Ile Asn Cys Met Glu Asn Asn Ser Gly Trp Lys Val Ser Leu
 130 135 140

-continued

Asn	Tyr	Gly	Glu	Ile	Ile	Trp	Thr	Leu	Gln	Asp	Thr	Gln	Glu	Ile	Lys
145					150					155					160
Gln	Arg	Val	Val	Phe	Lys	Tyr	Ser	Gln	Met	Ile	Asn	Ile	Ser	Asp	Tyr
				165					170					175	
Ile	Asn	Arg	Trp	Ile	Phe	Val	Thr	Ile	Thr	Asn	Asn	Arg	Leu	Asn	Asn
			180						185					190	
Ser	Lys	Ile	Tyr	Ile	Asn	Gly	Arg	Leu	Ile	Asp	Gln	Lys	Pro	Ile	Ser
		195					200					205			
Asn	Leu	Gly	Asn	Ile	His	Ala	Ser	Asn	Asn	Ile	Met	Phe	Lys	Leu	Asp
	210					215					220				
Gly	Cys	Arg	Asp	Thr	His	Arg	Tyr	Ile	Trp	Ile	Lys	Tyr	Phe	Asn	Leu
225					230					235					240
Phe	Asp	Lys	Glu	Leu	Asn	Glu	Lys	Glu	Ile	Lys	Asp	Leu	Tyr	Asp	Asn
			245						250					255	
Gln	Ser	Asn	Ser	Gly	Ile	Leu	Lys	Asp	Phe	Trp	Gly	Asp	Tyr	Leu	Gln
			260					265					270		
Tyr	Asp	Lys	Pro	Tyr	Tyr	Met	Leu	Asn	Leu	Tyr	Asp	Pro	Asn	Lys	Tyr
	275						280					285			
Val	Asp	Val	Asn	Asn	Val	Gly	Ile	Arg	Gly	Tyr	Met	Tyr	Leu	Lys	Gly
	290					295					300				
Pro	Arg	Gly	Ser	Val	Met	Thr	Thr	Asn	Ile	Tyr	Leu	Asn	Ser	Ser	Leu
305				310						315					320
Tyr	Arg	Gly	Thr	Lys	Phe	Ile	Ile	Lys	Lys	Tyr	Ala	Ser	Gly	Asn	Lys
			325						330					335	
Asp	Asn	Ile	Val	Arg	Asn	Asn	Asp	Arg	Val	Tyr	Ile	Asn	Val	Val	Val
			340					345					350		
Lys	Asn	Lys	Glu	Tyr	Arg	Leu	Ala	Thr	Asn	Ala	Ser	Gln	Ala	Gly	Val
		355				360						365			
Glu	Lys	Ile	Leu	Ser	Ala	Leu	Glu	Ile	Pro	Asp	Val	Gly	Asn	Leu	Ser
	370					375					380				
Gln	Val	Val	Val	Met	Lys	Ser	Lys	Asn	Asp	Gln	Gly	Ile	Thr	Asn	Lys
385				390						395					400
Cys	Lys	Met	Asn	Leu	Gln	Asp	Asn	Asn	Gly	Asn	Asp	Ile	Gly	Phe	Ile
		405							410					415	
Gly	Phe	His	Gln	Phe	Asn	Asn	Ile	Ala	Lys	Leu	Val	Ala	Ser	Asn	Trp
		420					425						430		
Tyr	Asn	Arg	Gln	Ile	Glu	Arg	Ser	Ser	Arg	Thr	Leu	Gly	Cys	Ser	Trp
	435						440					445			
Glu	Phe	Ile	Pro	Val	Asp	Asp	Gly	Trp	Gly	Glu	Arg	Pro	Leu		
	450					455				460					

(2) INFORMATION FOR SEQ ID NO:27:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 3891 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

- (ix) FEATURE:
 (A) NAME/KEY: CDS
 (B) LOCATION: 1..3888

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

ATG	CAA	TTT	GTT	AAT	AAA	CAA	TTT	TAT	AAA	GAT	CCT	GTA	AAT	GGT	
Met	Gln	Phe	Val	Asn	Lys	Gln	Phe	Asn	Tyr	Lys	Asp	Pro	Val	Asn	Gly
1					5				10					15	

-continued

TAA

3891

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1296 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

```

Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly
 1           5           10           15

Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Val Gly Gln Met Gln Pro
 20           25           30

Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg
 35           40           45

Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Pro Glu
 50           55           60

Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr
 65           70           75           80

Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu
 85           90           95

Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val
100           105           110

Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys
115           120           125

Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr
130           135           140

Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile
145           150           155           160

Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr
165           170           175

Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe
180           185           190

Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu
195           200           205

Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu
210           215           220

Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn
225           230           235           240

Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu
245           250           255

Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys
260           265           270

Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Tyr Asn
275           280           285

Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val
290           295           300

Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys
305           310           315           320

Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu
325           330           335

Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp
340           345           350

```

-continued

Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn
 355 360 365
 Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr
 370 375 380
 Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn
 385 390 395 400
 Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu
 405 410 415
 Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg
 420 425 430
 Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys
 435 440 445
 Ala Leu Asn Asp Leu Cys Ile Lys Val Asn Asn Trp Asp Leu Phe Phe
 450 455 460
 Ser Pro Ser Glu Asp Asn Phe Thr Asn Asp Leu Asn Lys Gly Glu Glu
 465 470 475 480
 Ile Thr Ser Asp Thr Asn Ile Glu Ala Ala Glu Glu Asn Ile Ser Leu
 485 490 495
 Asp Leu Ile Gln Gln Tyr Tyr Leu Thr Phe Asn Phe Asp Asn Glu Pro
 500 505 510
 Glu Asn Ile Ser Ile Glu Asn Leu Ser Ser Asp Ile Ile Gly Gln Leu
 515 520 525
 Glu Leu Met Pro Asn Ile Glu Arg Phe Pro Asn Gly Lys Lys Tyr Glu
 530 535 540
 Leu Asp Lys Tyr Thr Met Phe His Tyr Leu Arg Ala Gln Glu Phe Glu
 545 550 555 560
 His Gly Lys Ser Arg Ile Ala Leu Thr Asn Ser Val Asn Glu Ala Leu
 565 570 575
 Leu Asn Pro Ser Arg Val Tyr Thr Phe Phe Ser Ser Asp Tyr Val Lys
 580 585 590
 Lys Val Asn Lys Ala Thr Glu Ala Ala Met Phe Leu Gly Trp Val Glu
 595 600 605
 Gln Leu Val Tyr Asp Phe Thr Asp Glu Thr Ser Glu Val Ser Thr Thr
 610 615 620
 Asp Lys Ile Ala Asp Ile Thr Ile Ile Ile Pro Tyr Ile Gly Pro Ala
 625 630 635 640
 Leu Asn Ile Gly Asn Met Leu Tyr Lys Asp Asp Phe Val Gly Ala Leu
 645 650 655
 Ile Phe Ser Gly Ala Val Ile Leu Leu Glu Phe Ile Pro Glu Ile Ala
 660 665 670
 Ile Pro Val Leu Gly Thr Phe Ala Leu Val Ser Tyr Ile Ala Asn Lys
 675 680 685
 Val Leu Thr Val Gln Thr Ile Asp Asn Ala Leu Ser Lys Arg Asn Glu
 690 695 700
 Lys Trp Asp Glu Val Tyr Lys Tyr Ile Val Thr Asn Trp Leu Ala Lys
 705 710 715 720
 Val Asn Thr Gln Ile Asp Leu Ile Arg Lys Lys Met Lys Glu Ala Leu
 725 730 735
 Glu Asn Gln Ala Glu Ala Thr Lys Ala Ile Ile Asn Tyr Gln Tyr Asn
 740 745 750
 Gln Tyr Thr Glu Glu Glu Lys Asn Asn Ile Asn Phe Asn Ile Asp Asp
 755 760 765
 Leu Ser Ser Lys Leu Asn Glu Ser Ile Asn Lys Ala Met Ile Asn Ile

-continued

Gly Val Glu Lys Ile Leu Ser Ala Leu Glu Ile Pro Asp Val Gly Asn
 1205 1210 1215

Leu Ser Gln Val Val Val Met Lys Ser Lys Asn Asp Gln Gly Ile Thr
 1220 1225 1230

Asn Lys Cys Lys Met Asn Leu Gln Asp Asn Asn Gly Asn Asp Ile Gly
 1235 1240 1245

Phe Ile Gly Phe His Gln Phe Asn Asn Ile Ala Lys Leu Val Ala Ser
 1250 1255 1260

Asn Trp Tyr Asn Arg Gln Ile Glu Arg Ser Ser Arg Thr Leu Gly Cys
 1265 1270 1275 1280

Ser Trp Glu Phe Ile Pro Val Asp Asp Gly Trp Gly Glu Arg Pro Leu
 1285 1290 1295

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 12 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: Not Relevant
- (D) TOPOLOGY: Not Relevant

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 12
- (D) OTHER INFORMATION: /note= "The asparagine residue at this position contains an amide group."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

Cys Gln Thr Ile Asp Gly Lys Lys Tyr Tyr Phe Asn
 1 5 10

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: Not Relevant
- (D) TOPOLOGY: Not Relevant

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

His His His His His
 1 5

I claim:

1. A soluble fusion protein comprising a non-toxin protein sequence and a portion of the *Clostridium botulinum* type A toxin, said portion of the *Clostridium botulinum* type A toxin comprising a portion of the sequence of SEQ ID NO:28.
2. The fusion protein of claim 1, wherein said portion of the *Clostridium botulinum* type A toxin sequence comprises SEQ ID NO:23.
3. The fusion protein of claim 1, wherein said non-toxin protein sequence comprises a poly-histidine tract.
4. The fusion protein of claim 3, which comprises SEQ ID NO:26.
5. The fusion protein of claim 1, wherein said fusion protein is substantially endotoxin-free.
6. A host cell containing a recombinant expression vector, said vector encoding a protein comprising at least a portion of a *Clostridium botulinum* type A toxin protein sequence of

SEQ ID NO:28, and wherein said host cell is capable of expressing said protein as a soluble protein in said host cell at a level greater than or equal to 0.75% of the total cellular protein.

7. The host cell of claim 6, wherein said portion of a toxin comprises SEQ ID NO:23.

8. The host cell of claim 6, wherein said fusion protein comprises SEQ ID NO:26.

9. The host cell of claim 6, wherein said host cell is capable of expressing said protein in said host cell at a level greater than or equal to 20% of the total cellular protein.

10. A soluble fusion protein, comprising at least a portion of *Clostridium botulinum* C fragment linked to a poly-histidine tag.

* * * * *

-continued

Gly Val Glu Lys Ile Leu Ser Ala Leu Glu Ile Pro Asp Val Gly Asn	
1205	1210 1215
Leu Ser Gln Val Val Val Met Lys Ser Lys Asn Asp Gln Gly Ile Thr	
1220	1225 1230
Asn Lys Cys Lys Met Asn Leu Gln Asp Asn Asn Gly Asn Asp Ile Gly	
1235	1240 1245
Phe Ile Gly Phe His Gln Phe Asn Asn Ile Ala Lys Leu Val Ala Ser	
1250	1255 1260
Asn Trp Tyr Asn Arg Gln Ile Glu Arg Ser Ser Arg Thr Leu Gly Cys	
1265	1270 1275 1280
Ser Trp Glu Phe Ile Pro Val Asp Asp Gly Trp Gly Glu Arg Pro Leu	
1285	1290 1295

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 12 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: Not Relevant
- (D) TOPOLOGY: Not Relevant

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 12
- (D) OTHER INFORMATION: /note= "The asparagine residue at this position contains an amide group."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

Cys Gln Thr Ile Asp Gly Lys Lys Tyr Tyr Phe Asn
 1 5 10

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: Not Relevant
- (D) TOPOLOGY: Not Relevant

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

His His His His His
 1 5

I claim:

1. A soluble fusion protein comprising a non-toxin protein sequence and a portion of the *Clostridium botulinum* type A toxin, said portion of the *Clostridium botulinum* type A toxin comprising a portion of the sequence of SEQ ID NO:28.
2. The fusion protein of claim 1, wherein said portion of the *Clostridium botulinum* type A toxin sequence comprises SEQ ID NO:23.
3. The fusion protein of claim 1, wherein said non-toxin protein sequence comprises a poly-histidine tract.
4. The fusion protein of claim 3, which comprises SEQ ID NO:26.
5. The fusion protein of claim 1, wherein said fusion protein is substantially endotoxin-free.
6. A host cell containing a recombinant expression vector, said vector encoding a protein comprising at least a portion of a *Clostridium botulinum* type A toxin protein sequence of

SEQ ID NO:28, and wherein said host cell is capable of expressing said protein as a soluble protein in said host cell at a level greater than or equal to 0.75% of the total cellular protein.

7. The host cell of claim 6, wherein said portion of a toxin comprises SEQ ID NO:23.

8. The host cell of claim 6, wherein said fusion protein comprises SEQ ID NO:26.

9. The host cell of claim 6, wherein said host cell is capable of expressing said protein in said host cell at a level greater than or equal to 20% of the total cellular protein.

10. A soluble fusion protein, comprising at least a portion of *Clostridium botulinum* C fragment linked to a poly-histidine tag.

* * * * *